Exercise Two: Pie Charts

When to use a Pie Chart, when to use a bar chart and when to use both together!

The Packages required to complete this exercise have been loaded above. These include ggplot2, dplyr, DT, tools, RColorBrewer, readxl, reshape2, tidyR, scales, cowplot. We have also preloaded a function created to create a visualisation that utilizes the positives of both Bar and Pie Charts as seen in the Shiny Web App. This function is called, `pieBar` which takes the parameters:

1. Dataset <dataframe> - dataframe must be in the format [**name, proportion**]
2. Stacked <Bool> - If **TRUE**, your bar chart will be stacked, If **FALSE** it will be dodge.

As mentioned in the workshop, Pie Charts are really good if there are only 2 or 3 wedges, however if you have more wedges it is harder to visually disern exact information and to make comparisions. There are three datasets provided below, take a look the data and determine which visualisation is most appropriate.

Example One

The data below is stored as `DatasetOne` and the data is taken from https://countrymeters.info/en/Australia.

```
##       Sex Percentage
## 1 Female      0.498
## 2   Male      0.502
```

What do you think would be the most appropriate way to visualise this information? Fill in the missing parameters below. Don't forget to use the ggplot2 **cheat sheet** to help you.

```
#ggplot(data = DatasetOne, aes(x = "", y = Percentage, fill = Sex)) +
#_____(stat = "identity") + _____ (_____, start=____)
```

Don't forget to uncomment the function in order to test your answer.

ANSWER: If you take a look at the dataset, you can note that if we used a Pie Chart there would only be two wedges. Therefore, if you plotted a pie chart you would have been correct. Note, if you plotted a bar chart that would also have been correct

To explore, lets take a look at how easy it is to compare two countries population distributions using pie charts. The Pie Charts below compare Australia to Saudi Arabia.

```
## Coordinate system already present. Adding new coordinate system, which will replace the existing one
## Coordinate system already present. Adding new coordinate system, which will replace the existing one
```
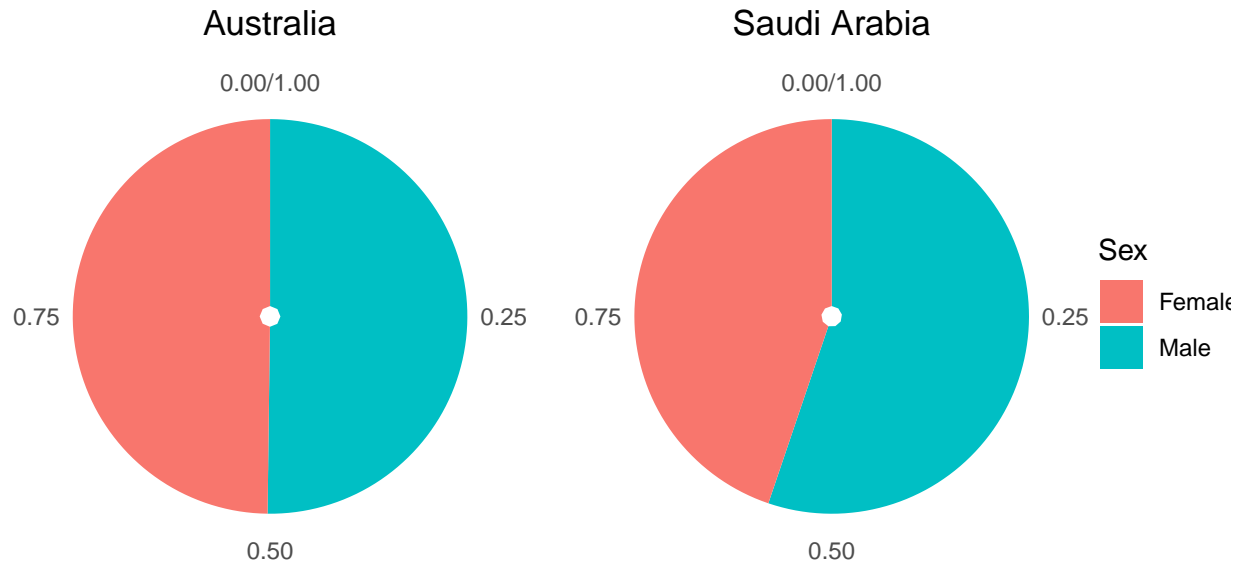
Example Two

Dataset Two, stored in `DatasetTwo`, contains the proportion of the population of Australia in each 5 year age group. The data was taken from Australian Bureau of Statistics, Census of Population and Housing 2011 and 2016. Compiled and presented by .id , the population experts (https://profile.id.com.au/australia/five-year-age-groups).

```
## # A tibble: 18 x 4
##    age_groups   Number percentage Capital_Cities
##    <chr>         <dbl>      <dbl>          <dbl>
##  1 0 to 4      1464776       6.26           6.40
##  2 5 to 9      1502644       6.42           6.38
##  3 10 to 14    1397182       5.97           5.85
##  4 15 to 19    1421612       6.07           6.09
##  5 20 to 24    1566792       6.70           7.19
##  6 25 to 29    1664609       7.11           7.78
##  7 30 to 34    1703852       7.28           7.96
##  8 35 to 39    1561686       6.67           7.14
##  9 40 to 44    1583254       6.77           6.98
## 10 45 to 49    1581460       6.76           6.80
## 11 50 to 54    1523557       6.51           6.34
## 12 55 to 59    1454332       6.21           5.85
## 13 60 to 64    1299406       5.55           5.09
## 14 65 to 69    1188989       5.08           4.54
## 15 70 to 74     887721       3.79           3.35
## 16 75 to 79     652671       2.79           2.50
## 17 80 to 84     460555       1.97           1.80
## 18 85 and over  486847       2.08           1.96
```

Taking a look at the data table above, what visualisation do you think is the most approprate. I've set up the barebones of the R code below. Fill in the blanks below or create your own visualisation from scratch.

```
#ggplot(data = DatasetTwo, aes(x = _____ , y = percentage, fill = age_groups)) +
#_____(position ="_____", stat = "identity") + theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

ANSWER: Taking a look at the data, we can see that if we were to plot the data using a pie chart we would have 18 wedges, which is more than the prescribed three wedges. Therefore, we should plot the data as a bar chart. The question is, What type of bar chart would be best? Stacked or Dodge. Which did you choose?

Lets take a look at both of these to determine which one displays the data the best.

Dodge Bar Chart vs Stacked Bar Chart
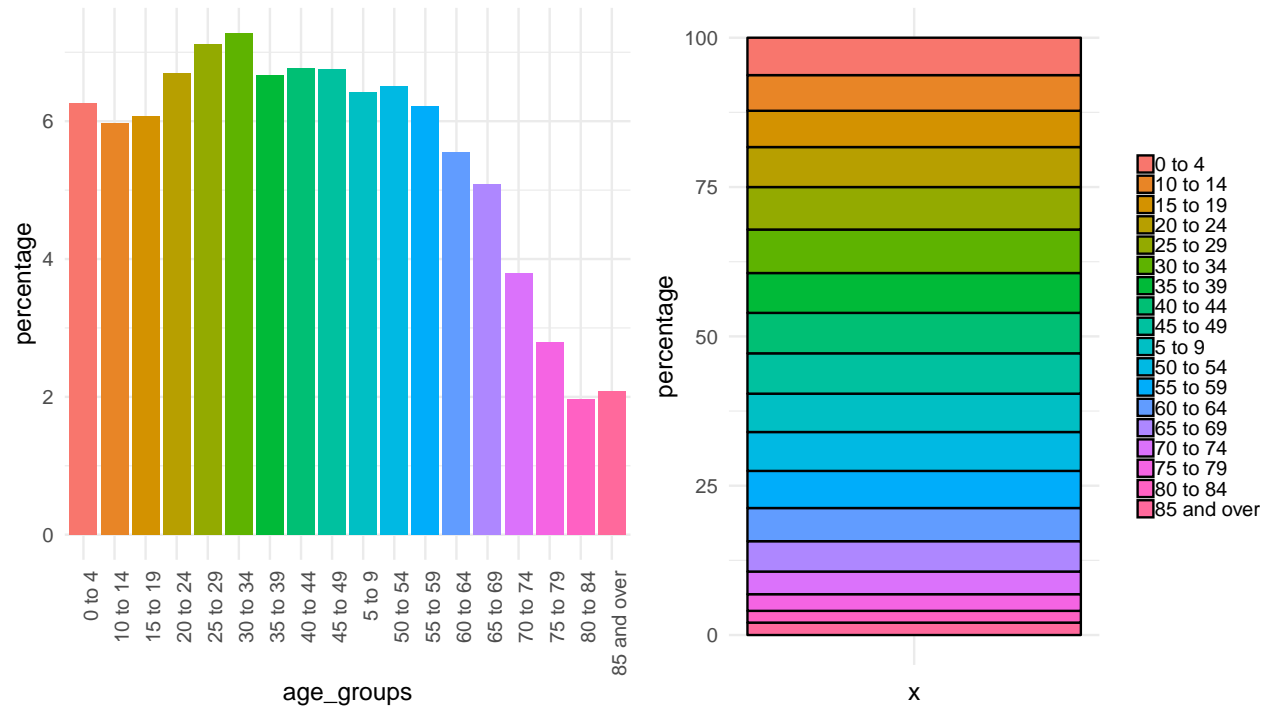
Taking a look at these two visualisations, it appears easier to determine inidiviual values using the 'Dodge' bar chart instead of the stacked.

Exercise Three

Dataset Three stored as `DatasetThree`, is the percentage of Australians that list each country as their country of birth in 2011. The data was taken from http://quickstats.censusdata.abs.gov.au/census_services/getproduct/census/2011/quickstat/0.

```
##      birthCountrie Percentage
## 1        Australia      66.70
## 2   United Kingdom       4.64
## 3      New Zealand       2.21
## 4            China       3.09
## 5            India       2.66
## 6      Phillippines       1.19
## 7          Vietnam       1.36
## 8            Italy       0.96
## 9     South Africa       0.82
## 10        Malaysia       0.82
## 11       Sri Lanka       0.65
## 12         Germany       0.40
## 13           Other      15.00
```

Taking a look at the data, we can notice three points that can help you to determine what is the appropriate visualisation.
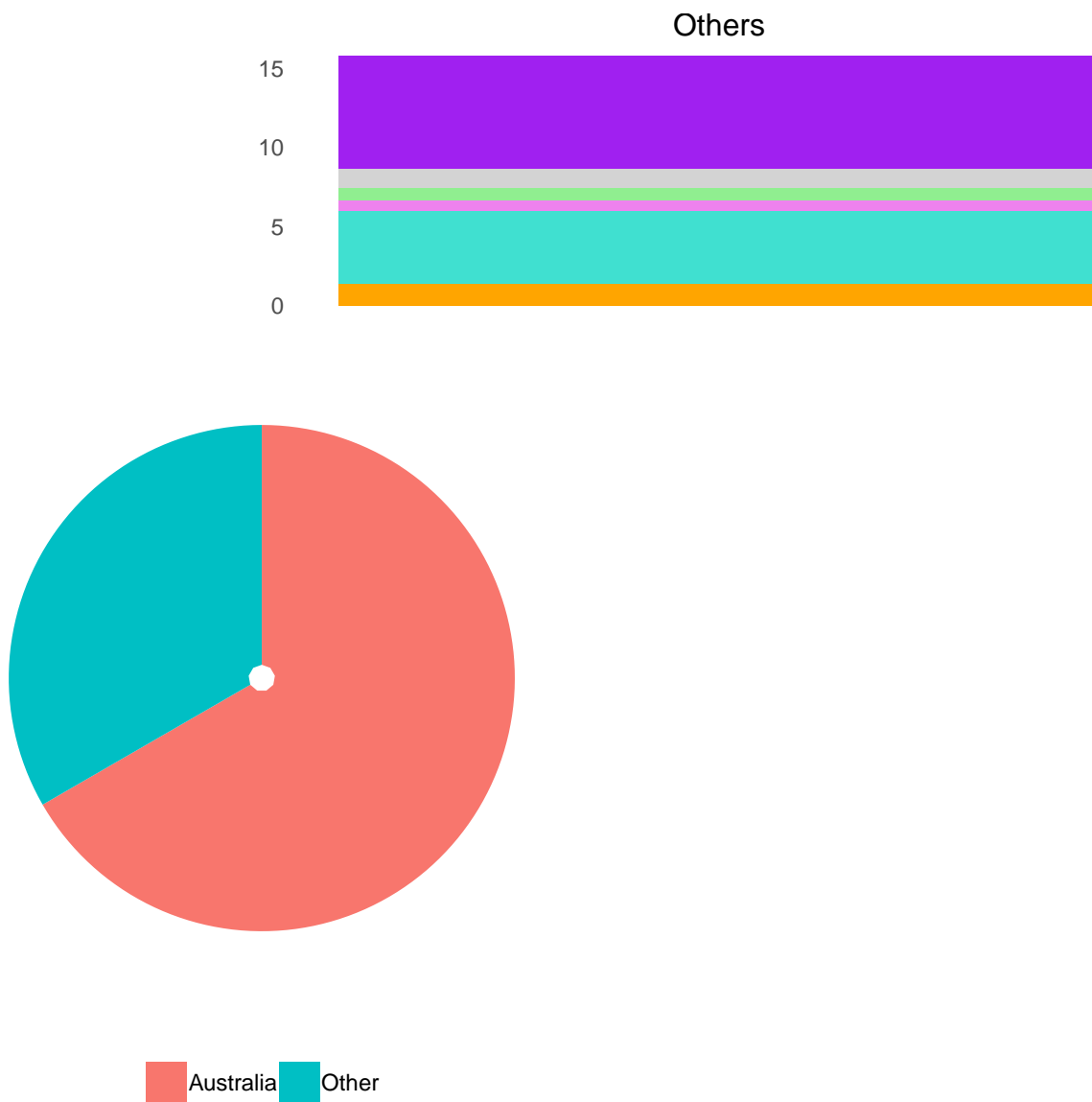
1. See that a large proportion, **66.7%** of Australians were in fact born in Australia.
2. None of the other countries percentages come anywhere close to this percentage.
3. There are 13 wedges.

These three conditions provide us with a perfect example to use the `pieBar` function. Use the `pieBar` function below to create this new visualisation with the data.
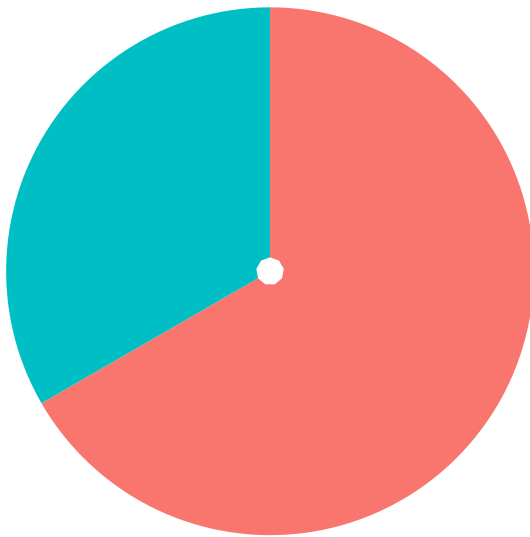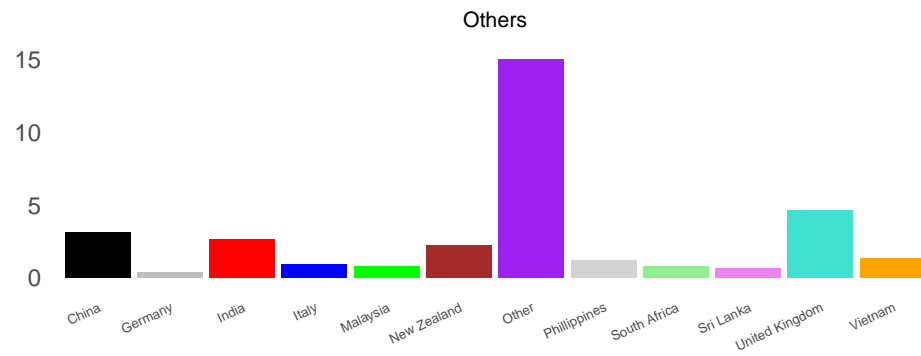
```
#pieBar(Dataset = '____' , Stacked = '____')
```

ANSWER: Does your function look the same as below? Did you use a stacked or dodge bar chart and why?

```
pieBar(Dataset = DatasetThree, Stacked = TRUE)
```

```
pieBar(Dataset = DatasetThree, Stacked = FALSE)
```

Others

Australia   Other